4 EXCESSIVE PROOFREADING BY T7 DNA POLYMERASE

Abstract

DNA polymerase catalyzes the accurate transfer of genetic information from one generation to the next, and it is thus of vital importance that replication is faithful. DNA polymerase fulfills the strict requirements on fidelity by a combination of mechanisms: (i) high selectivity for the correct nucleotide to be incorporated, (ii) slowing down of the replication rate after misincorporation, and (iii) proofreading activity by excision of the misincorporated base. To elucidate the kinetic interplay between replication and proofreading we here use high-resolution optical tweezers to probe how the DNA duplex stability affects replication by bacteriophage T7 DNA polymerase. Our data show highly irregular replication dynamics, with frequent pauses and directional changes as the polymerase cycles through the states that govern the mechanochemistry behind high-fidelity T7 DNA replication. We construct a kinetic model that incorporates existing biochemical data as well as the novel states we observe. We fit the model directly to the acquired pause-time and run-time distributions. Our findings indicate that the main pathway for error correction is by exchange: DNA polymerase unbinds from the polymerase active site, followed by biased binding into the exonuclease active site. Curiously, this proofreading mechanism removes far more bases than erroneous bases expected to be incorporated; this appears to be the price bacteriophage T7 pays for the high-fidelity replication of its genome.

4.1 Introduction

Life depends on the accurate transfer of genetic information between generations. With genomes consisting of up to several gigabase pairs [60], it is essential that DNA polymerase (DNAp), the principal enzyme catalyzing replication of DNA, works with a very high fidelity. Part of the selectivity for incorporating correct—rather than the incorrect—nucleotides results from the compounded effects of nucleotide discrimination and the slow incorporation of erroneous bases into the nascent DNA strand. This selectivity limits the number of errors incorporated into the product to $10^{-5} - 10^{-7}$ per base pair [85,98,138]. For the 40 kb genome of bacteriophage T7 this would mean about 0.04 mutations per replication. Erroneous nucleotides that do get incorporated into the nascent strand cannot form correct Watson-Crick base pairs with the template, and therefore destabilize the primer-template structure (PTS) at the polymerase active site (*pol*) [86]. For *family A* DNAp enzymes (such as T7 DNAp), this triggers the transfer of the 3'-end of the nascent DNA strand from (*pol*) to a second active site: the exonuclease active site (*exo*). In *exo*, the 3'-end is actively digested, removing erroneously incorporated bases [99]. This post-incorporation proofreading mechanism brings the error rate down two additional orders of magnitude, to $10^{-7} - 10^{-9}$ [179].

It has been shown that T7 DNAp, when starved of dNTPs, processively removes bases from dsDNA while its *exo* active site interacts with DNA [42, 94]. In the presence of dNTPs, processive exonucleolysis by DNAp has not been observed. The bulk approaches applied in these studies, however, only observe the time-averaged activity and processive exonucleolysis runs could be obscured by more frequent runs of polymerization activity. Using bulk methods, nucleotide incorporation kinetics has been measured on an exonuclease-deficient mutant of T7 DNAp, to make sure that potential *exo* activity did not reduce the measured polymerization rate [138]. Consequently, it has not been possible to directly quantify the dynamics between polymerization and exonucleolysis in bulk experiments.

Single-molecule force-spectroscopy experiments, on the other hand, do not require time or ensemble averaging. Such experiments have been performed to elucidate the mechanochemistry of DNAp, using wild-type DNAp to study *pol* and *exo* activity simultaneously and the kinetic balance between them. Tension on the template strand of the DNA destabilizes the PTS and triggers transfer of the DNA 3'-end to the *exo* site. The effect of force on the PTS is similar to that of the incorporation of an erroneous base, and has been used to study exonucleolysis by DNA polymerase [79, 180]. It has previously been shown that the apparent rate of polymerization goes down as tension increases up to ∼35 pN [180], after which DNAp seems to exclusively perform exonucleolysis [79, 115].

In a single-molecule FRET study, using fluorescent base analogues, a related *family A* polymerase (Klenow fragment) has been shown to bind relaxed, matched PTS with both its *pol* and *exo* active sites [28,35]. Similar data has been observed for for a DNAp active in translesion [18], and with carcinogenic adducts that induced binding into the *exo* active site [173]. These studies raise the question of whether other polymerases also bind and process DNA in *exo*, even with an unperturbed correctly base-paired PTS. Any

short-lived *exo* activity could have been missed in previous studies due to the limited spatial and temporal resolution. Importantly, a hitherto unresolved *exo* activity, in the absence of a destabilized PTS, would demonstrate that replication is highly stochastic also in the absence of incorporated errors.

The replicative T4 DNA polymerase (*family B*) was shown to preferentially perform processive exonucleolysis when polymerization was obstructed by dsDNA ahead [117]. Using magnetic tweezers, a hairpin ahead of T4 DNAp was perturbed to varying extent. The regression pressure of the fork at low tensions inhibited forward motion of DNAp. Consequently, the equilibrium shifted towards processive exonucleolysis. T7 DNAp showed a similar behavior in this study: exonucleolysis was observed at low tensions, when the dsDNA restrained the DNAp from polymerization.

Here we study the kinetics of T7 DNA polymerase using high-resolution optical tweezers. Our results show several previously unresolved kinetic states of DNAp, and provide direct access to probabilities and rates of the transitions between them. We show that both *pol* and *exo* activities coexist over the full range of tensions examined. Moreover, we show that the apparent slowdown with tension in net replication rate results from a gradually increased bias for binding into *exo* over *pol* due to tension-induced destabilization of the PTS. Based on our high-resolution data, we propose a kinetic model for T7 DNAp that differs substantially from previous schemes [42,85]. Our analysis challenges the notion of DNAp proofreading as a precisely controlled mechanism that engages only when an error is encountered [4]. Instead, replication appears prone to false positives (removal of correctly incorporated bases) as a means to ensure the elimination of false negatives (erroneously incorporated bases that are not removed). This "paranoia" of DNAp might be a general evolutionary strategy employed by high-fidelity molecular machines to deal with thermal noise at the smallest scales of living systems.

4.2 Results

To investigate the kinetic balance between replication and proofreading by T7 DNAp we used a high-resolution double-trap optical tweezers assay that allows direct observation of polymerization and exonucleolytic activity [180]. A double-stranded DNA (dsDNA) molecule of 8 kb with a 25-nucleotide (nt) 5'-overhang is attached between two optically trapped polystyrene microspheres (Figure 4.1A). Because the end-to-end lengths of ds- and single-stranded DNA (ssDNA) differ significantly at tensions above ~6 pN, the conversion of ss- to dsDNA by the replication reaction results in a decreased length of the DNA tether, while exonucleolysis results in tether lengthening (Figure 4.1B). Using the DNA tether length and the known tension, the position of the ss-dsDNA junction and the DNAp can be calculated. Our current instrument has a better time (events as short as 0.4 sec can be detected) and spatial (20 bps at 45 pN, Figure S4.1) resolution than previously used instruments. To ensure the highest resolution, we have only applied forces above the crossover point (~6 pN). This allows us to study the different kinetic states of T7 DNAp in great detail. In Figure 4.1C a representative trace of T7 DNAp activity is presented, showing bursts of polymerization and exonucleolysis activity, in-

terspersed by pauses. The traces are analyzed by determining the transition points between trend changes (see Appendix) and thereby segmenting the traces in periods of distinct activities (polymerase, exonucleolysis and four kinds of pauses) (Figure 4.1D). From the slope of bursts of polymerization (n=549) and exonucleolysis (n=4,362) activity, their rates can be directly determined (Figure S4.2). This analysis is performed on traces of T7 DNAp activity performed at different DNAp concentration and tensions on the DNA template.

4.2.1 Polymerization and exonucleolysis at different tensions

We first looked at the *pol* and *exo* rates as a function of DNA tension during bursts of activity (Figure 4.2A). In accordance with previous studies we observe that the rate of polymerization decreases with DNA tension up to 35 pN. The polymerization rates we find—up to 500 nt s^{-1} at a tension of 15 pN—are significantly higher than reported before using bulk (300 nt s^{-1}) and a single-molecule (100 nt s^{-1}) experiments [179,180]. Our results are, however, in close comparison to the rates reported for polymerization in a magnetic tweezers experiment [117]. The difference with the seminal study of T7 DNAp in an optical tweezers experiment is, most likely, a consequence of our improved temporal resolution, which allows identification (and exclusion) of short—previously undetected—pauses in between replication events [180]. Surprisingly, we now also observe *pol* activity at high tension, even at tensions close to the dsDNA overstretching transition at 65 pN [62]. The polymerization rate in these particular events is relatively low (∼80 nt s^{-1}) and does not appear to depend on a further increase in tension. We observe exonucleolysis at a constant rate of 145 ± 1 nt s^{-1} (mean ± SEM, n=3990) over the full range of tensions studied. Previously, processive force-induced exonucleolysis was only reported at tensions above 35 pN, at a rate of ∼25 nt s^{-1}. The rate is in the same order as the rate found with magnetic tweezers (∼200 nt s^{-1}) [117]. Next, we determined the time DNAp stays active in both *pol* and *exo*. At each tension, the duration of polymerization and exonucleolysis events are exponentially distributed (Figure S4.3). Fits to the durations yield an off rate of ∼1 s^{-1} for both *pol* and *exo* activity that does not depend strongly on DNA tension (inset Figure 4.2A). For *pol* this value is consistent with rates found in bulk experiments, while for *exo* it is three orders of magnitude slower than previously reported (1,000 s^{-1}) [42]. It should be noted, however, that the previously reported *exo* rate is not the result of a measurement on dsDNA, but is inferred from the rate of exonucleolysis on ssDNA substrates.

To determine the probability of either *pol* or *exo* activity occurring, we counted the number of *exo* events and divided this by the total number of *pol* and *exo* events per tension bin. Figure 4.2B shows that the probability of binding into *exo* increases drastically above 35 pN. This has consequences for fidelity: incorporation of erroneous bases destabilizes the PTS and triggers their removal by predominant binding into *exo*. In contrast to our observation of *exo* activity also at small tensions, previous studies reported no *exo* activity below 35 pN [79,180]. We believe that this apparent discrepancy with our data is the result of the better temporal and spatial resolution of our instrument. This is supported by the fact that the net polymerization / exonucleolysis rates at

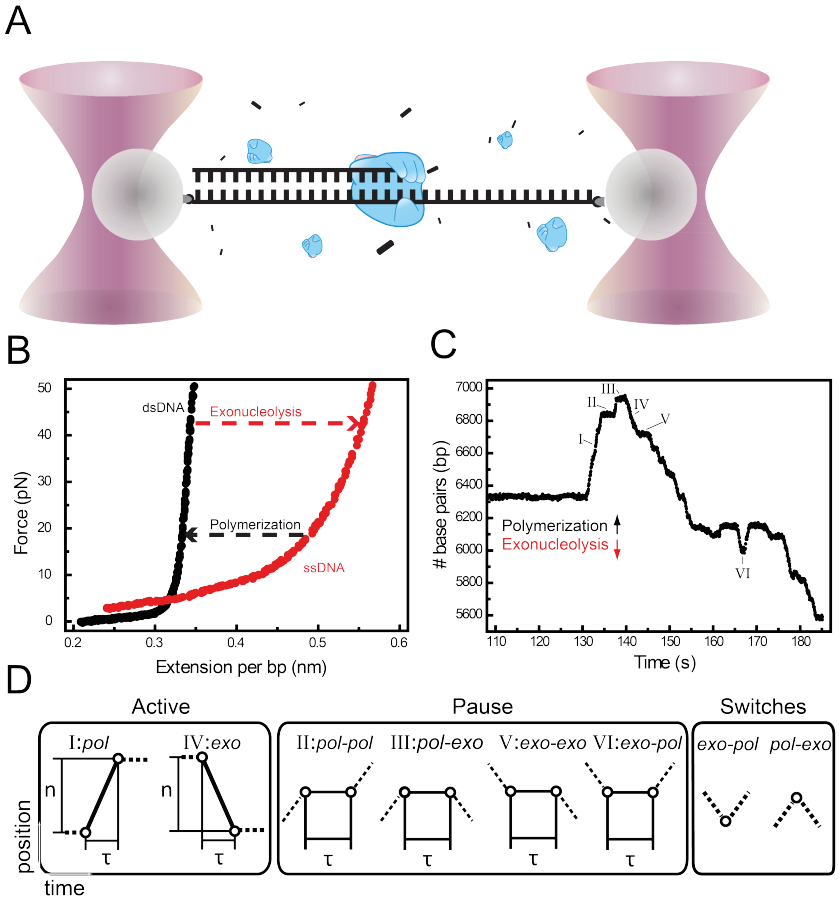

Figure 4.1
Optical tweezers assay to monitor T7 DNA polymerase activity. (A) Schematic of experimental layout. A DNA molecule with a 5'-overhang (right) is held between two optically trapped microspheres. In the presence of T7 DNA polymerase and nucleotides, activity is measured by determining DNA length changes while clamping the force. (B) Force-extension curves of dsDNA and ssDNA. At tensions above 6 pN, polymerization by DNA polymerase at constant tension results in shortening of the DNA construct, while exonucleolysis leads to lengthening (dashed arrows). (C) Typical time trace of T7 DNA polymerase activity at a force of 45 pN, with six different behaviors indicated. (D) Schematic of observed behaviors: polymerase, exonucleolysis, four different types of pauses and two direct switches in activity. In these diagrams, solid lines between circles represent current behavior while dashed lines indicate preceding and subsequent behavior. The axes are equivalent to panel C.

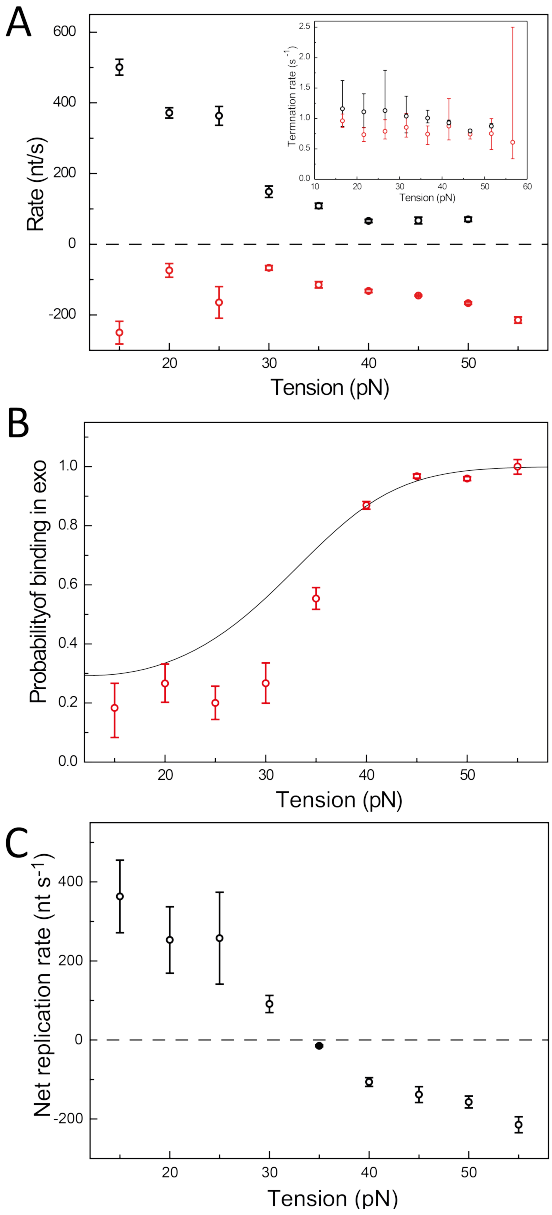

Figure 4.2
Polymerization and exonucleolysis occur over the full range of tensions applied to the DNA. The rate of polymerization (black) decreases with tension, while that of exonucleolysis is relatively constant (red; 145 ± 1 nt s^{-1} mean ± SEM, n=3990). Inset: off rates of both polymerization (black) and exonucleolysis (red) activities do not depend on tension. (B) The relative probability of DNAp binding with its *exo* active site depends strongly on tension (mean ± SD). The solid line is *not* a direct fit to this data, but is calculated using the rates fitted to the complete time statistics of pauses (see Appendix). (C) Reconstruction of the net replication rate, obtained by adding the polymerization and exonucleolysis rates weighted for their relative occurrences showing a switch in net activity from polymerization to exonucleolysis at 30 pN.

different tensions (attained by adding *pol* and *exo* rates weighted with their relative occurrences) yields values very similar to those obtained in previous, lower-resolution experiments (Figure 4.2C) [79, 180].

4.2.2 Pausing reveals multiple kinetic states

While the biological functions of *pol* and *exo* in ensuring proper DNA replication are straightforward, the role of pausing remains unclear. A pause either indicates that DNAp is not present at the ss-dsDNA junction, or present and not actively bound, preventing catalysis by another DNAp. To unravel the origin of replication pauses, we systematically investigated the pause statistics and how they depend on tension and DNAp concentration. With our high-resolution experiments, we cannot only distinguish pauses that occur during transitions between *pol* and *exo*, but we are also able detect pauses during *pol* and *exo* (Figure 4.1D).

First, we considered the distribution of pause durations. This distribution is double-exponential with two well-separated characteristic times (Figure 4.3A). Hence, the pauses do not represent a single state, but a composite of at least two distinct states. When comparing different concentrations of DNAp, only the shorter pauses are found to depend significantly on concentration (inset Figure 4.3A). The shorter pauses most likely represent the binding of DNAp from solution to the DNA with either *pol* or *exo*, since these processes are concentration-dependent. Long pauses, instead, most likely are caused by a state in which DNAp is bound to the DNA, but replication is blocked, since their average duration is DNAp-concentration independent. Blocked replication could be caused by DNAp being bound tightly to ssDNA close to the junction, or bound to the junction in a wrong configuration. We further observe that duration and relative occurrence of long pauses do not significantly depend on prior activity, DNAp concentration and tension on the DNA.

4.2.3 An additional pause state during polymerization

More insight into the paused states is obtained by considering the statistics of activities that precede and follow pauses. Based on existing models [42], where exiting activity from both *exo* and *pol* takes the system to the same solution state, we would expect the statistics of binding into *exo* and *pol* to be independent of the activity prior to the pause. At low tensions this is indeed what we observe (Figure 4.3B). Surprisingly, when the PTS is destabilized by tensions of 30 pN or higher, the system retains a memory of the previous activity (Figure 4.3B): DNAp enzymes paused during *pol* activity have a higher probability of resuming *pol* activity than proteins paused during *exo* activity. We hypothesize that the memory effect is due to an inactive state accessible only from *pol*, in which DNAp does not unbind from the DNA and thus retains a memory of where it originated. To account for the memory effect, the additional paused state has to be accessible from *pol* and return to *pol* with high probability. While this model intuitively explains the memory effect, we have considered alternative models,

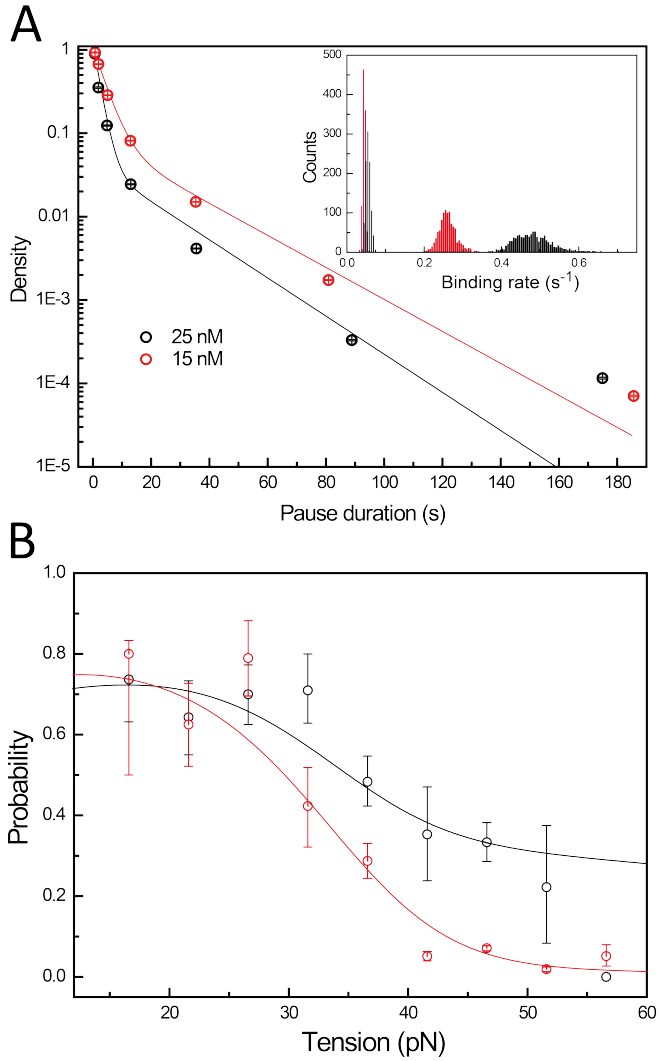

Figure 4.3
Distribution of pause durations. (A) Black: high concentration (25 nM, 1535 pauses); red: low concentration (15 nM, 1430 pauses); solid lines: double-exponential fits. For clarity, data and fits have been normalized to 1 at 0.4s. Inset: histograms of binding rates for both concentrations (1000 bootstraps). (B) The probability of binding in the pol active site after a pause that has been preceded with polymerization (black) or exonucleolysis activity (red) (mean ± SD). The solid lines are *not* a direct fit to this data, but are calculated using the rates fitted to the complete time statistics of pauses (see Appendix).

but for reasons discussed in Appendix they have been discarded. At high tensions this additional pause state becomes accessible from *pol*. In our model, this *pol* pause could e.g. be part of the DNAp incorporation cycle: the state might be entered when DNAp checks the base pairing of the incoming nucleotide with the template strand, but is unable to incorporate it instantly because of tension-induced deformation of the PTS. DNAp could escape from this paused state either by release of the incoming nucleotide or substantially slowed down addition to the nascent strand.

4.2.4 *Exo* to *pol* direct switching

We also investigated the possibility of so-called direct switches, fast transitions from *exo* to *pol* or *pol* to *exo* activities, without DNAp unbinding from the PTS. Obviously, pause durations shorter than our time resolution cannot be measured, while a switch in enzyme activity can be inferred from the change in the direction of the trace. We observe direct switches, but our time resolution is limited. This means that these apparent direct switches might (in part) be due to pauses too short to be resolved given the time resolution of our instrument. To estimate the fraction of switches that originate in sub-resolution solution pauses—rather than being direct switches—we extrapolate our fits of the solution pauses into the sub-resolution regime to calculate the number of expected sub-resolution pauses. The expected number of sub-resolution pauses going from *pol* to *exo* was 19 compared, to 8 detected apparent direct switches. Therefore, we find no evidence for a direct switch from *pol* to *exo*. In contrast, for *exo* to *pol* 20 sub-resolution pauses were expected, far less than the 84 apparent direct switches detected. Taken together, we have clear evidence for a direct switch only *exo* to *pol* (Figure S4.4). In close agreement with our findings is the study of T4 DNA polymerase [117]. By applying varying destabilizing tensions to a downstream DNA hairpin, high-resolution data was obtained on polymerization, exonucleolysis, pauses and their dynamics. Interestingly, the transition from *exo* to *pol* is found to occur instantly ∼90% of the time, whereas the transition from *pol* to *exo* is interspersed by a pause.

4.3 Kinetic model T7 DNAp: replication, proofreading and pausing

Previously, a three-state model for T7 DNAp activity was inferred from biochemical studies [42]. For clarity, we start with this model (illustrated in black in Figure 4.4A), and expand it step by step to account for our new experimental observations. The original model connects processive polymerase activity (P-activity) and exonuclease activity (E-activity) via a state in which DNAp is free in solution (S-pause). In addition to a changeover from *exo* to *pol* activity through DNAp unbinding and binding, previous studies have obtained evidence for direct *exo* to *pol* switching, and inferred that the reverse transition, switching from *pol* to *exo* also occurs [42].

With our high-resolution data and ability to track the kinetic history of individual enzymes, we have demonstrated that this kinetic model does not describe our data as it does not explain the observed memory of the state prior to the pause. Moreover, the observed long-pause state does not fit in the 3-state system. Hence, we propose three modifications of the model (Figure 4.4A). First, we observed evidence for direct *exo* to *pol* switching but not for direct *pol* to *exo* switching. Therefore we omit the latter transition in our model. Second, an additional pause state is added to account for the long pauses (see Figure 4.3A). We have not observed any indication that these pauses occur predominantly before or after *pol* and *exo* activity, and their occurrence does not drastically depend on DNAp concentration. Therefore, the "L-pause" (illustrated in green) can be accessed directly by DNAp binding from solution to the DNA. Furthermore, the L-pause duration is governed by a concentration-independent unbinding rate. Third, a pause state that might be entered during incorporation is added (I, illustrated in blue) to account for the memory effect DNAp displays at high tension. A comparable model was obtained for T4 DNAp in a study with magnetic tweezers [117]. There, it was shown that an intermediate paused state separates the *pol* and *exo* active sites. In addition, a paused state during polymerization and a direct transition from *exo* to *pol* were revealed, similar to the model we present here. We have considered three alternative models that in principle could account for our observations. These model are discussed in the Appendix: a completely interconnected 4-state model, a model in which fraying of the DNA accounts for the memory effect, and a model that has a paused state branching from *exo*. As discussed, all three models can be excluded for different reasons.

On basis of our new kinetic model, we calculated the theoretically expected distributions of the activation times for all types of detectable transitions (see Appendix). We then used maximum-likelihood estimation with respect to the theoretically derived distributions to extract the rate constants. To fit the model without making assumptions for tension dependence, the data was binned into three tension bins. We further bootstrapped the data 300 times to estimate the sensitivity in fit parameters. The results of the maximum-likelihood estimates are shown in Figure 4.4B-E, together with the fits to a simple force-dependent rate of the form:

$$\boldsymbol{k}_{XY} = \boldsymbol{k}^0_{XY} e^{-\delta_{XY} f / k_B T}$$

(for $\boldsymbol{k}_{SP}$ a more complicated form was used, see below). Consistent with our direct observations, the binding rate from solution to the *pol* active site ($\boldsymbol{k}_{SP}$) decreases with tension, while the binding rate from solution to the *exo* active site ($\boldsymbol{k}_{SE}$) increases. At the same time, the binding rate into the long pause appears to be insensitive to tension (Figure 4.4B). The binding rate into *pol* ($\boldsymbol{k}_{SP}$) is the only rate that is not fitted well by the simple force-dependence shown above. In this case we expand the exponent to second order in the force, and fit the higher order exponential function:

$$\boldsymbol{k}_{SP} = \boldsymbol{k}^0_{SP} e^{-(\delta_{SP} f + \gamma_{SP}(f-\bar{f})^2 - \gamma_{SP}\bar{f}^2)/k_B T}$$

Here we have included the average force $\bar{f} = 33.1$ pN of the experiments in the second order force term to maintain the interpretation of δ_{SP} as a distance to a transition state in the experimentally probed range of forces.

Direct switching from *exo* to *pol* activity predominantly occurs at low tension ($\boldsymbol{k}_{EP}$), while all other rates out of *pol* and *exo* activities ($\boldsymbol{k}_{PS}$, $\boldsymbol{k}_{PI}$, $\boldsymbol{k}_{ES}$) are largely insensitive to tension (Figure 4.4C). The probability of entering an incorporation pause during a single nucleotide incorporation event ($\boldsymbol{k}_{PI}/\boldsymbol{k}_{pol}$) increases with tension, since the polymerization rate decreases (Figure 4.4D). We speculate that incorporation pauses are caused by a distorted PTS. Over the full range of tensions tested, the escape rate from this incorporation pause back to *pol* activity ($\boldsymbol{k}_{IP}$) is of similar magnitude as the binding rates from solution ($\boldsymbol{k}_{SP}$ and $\boldsymbol{k}_{SE}$) (Figure 4.4E). The escape rate from the long-pause state to solution ($\boldsymbol{k}_{LS}$) also remains constant and is considerably lower than all other rates. From the fits we can also estimate the distances (DNA length changes) to the transition states for all transitions (see Appendix). The DNA length changes are all in the sub-nanometer range. For example, the DNA length change occurring between *pol* and *exo* activity is consistent with that of the melting of ~3 base pairs (0.62 nm), in agreement with structural data [9, 43]. Moreover, we can use the obtained rates to make predictions on the probability of binding *exo* (Figure 4.2B), and the memory effect (Figure 4.3B). These predictions are plotted in the figures (solid lines) and show that the obtained rates fit our data very well. In conclusion, our kinetic model reveals new pathways in the mechanochemistry of T7 DNAp and is consistent with our single-molecule data, biochemical studies [42, 179], as well as structural studies [9, 43].

4.4 Discussion

Using high-resolution optical tweezers we have examined how the dynamics of DNAp change with the amount of tension destabilizing the PTS. Since our data has sufficient resolution to discriminate between different dynamical behaviors—keeping track of the origin, destinations, and transition times for each transition—we are able to form a clear picture of the underlying dynamical process. In short, we find evidence for a bound inactive state, an incorporation pause induced by a destabilized PTS, direct switching from *exo* to *pol* activities; we find no evidence for *pol* to *exo* direct switching. Here we focused on the effects of destabilizing the PTS, but also probed a sufficient concentration range of DNAp to detect concentration dependencies.

In our experiments we observe both a slowdown of polymerization and an increased probability of being trapped in the incorporation pauses at increasing tension. The decrease of the polymerization rate can be explained by T7 DNAp undergoing force-sensitive structural changes during polymerization, when the "finger" domains have to close to correctly align and incorporate a nucleotide [148]. In a physiological setting, our high-tension situation corresponds to an erroneous nucleotide, which cannot properly base pair with the template strand, increasing both the probability of unbinding of the incoming nucleotide and the probability of unbinding of DNAp from its *pol* active site and rebinding with the *exo* active site [35, 88, 165]. We find that the probability

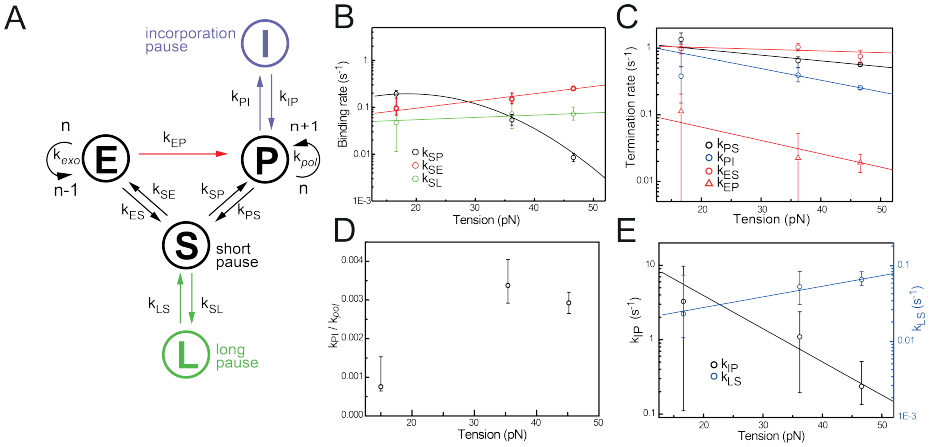

Figure 4.4
Kinetic model of T7 DNA polymerase fitted to the experimental data. A: Scheme of kinetic model with rates indicated. Original model (based on biochemical data [42]) is indicated in black. On basis of our experimental results three modifications have been made. Green: long pause state; blue: incorporation pause; red: direct switching only occurs from exo to pol active site. (B)–(E) Results of maximum-likelihood fitting of the model in (A) to the experimental data. Sensitivity in the parameters is estimated by bootstrapping the data (300 times) and the 95% confidence interval is shown. Except for $\boldsymbol{k}_{SP}$, all rates are fitted with $\boldsymbol{k}_{XY} = \boldsymbol{k}^0_{XY} e^{-\delta_{XY} f / k_B T}$. $\boldsymbol{k}_{SP}$ is fit to a second-order force dependent rate. (B) Binding rates ($\boldsymbol{k}_{SP}$, $\boldsymbol{k}_{SE}$, $\boldsymbol{k}_{SL}$) out of solution as function of tension. (C) Off rates of polymerization and exonucleolysis activities ($\boldsymbol{k}_{PS}$, $\boldsymbol{k}_{PI}$, $\boldsymbol{k}_{ES}$) and direct-switching rate from *exo* to *pol* ($\boldsymbol{k}_{EP}$) as a function of tension. (D) Ratio of the rate into incorporation pause ($\boldsymbol{k}_{IP}$) over the polymerization rate ($\boldsymbol{k}_{pol}$) as a function of tension. (E) The recovery rates from incorporation pauses ($\boldsymbol{k}_{PI}$) and long pauses ($\boldsymbol{k}_{LS}$) as a function of tension.

of entering an incorporation pause increases fivefold when the tension is increased to over 35 pN. Consequently, these pauses compete with polymerization and lower the net polymerization rate of bound DNAp. It is remarkable that polymerization still occurs at tensions as large as 50 pN, when the PTS is perturbed. This relative indifference of polymerase activity for PTS perturbation might help the enzyme to overcome other obstacles during replication, such as proteins bound to the ssDNA (gp2.5) or DNA hairpins.

We observe *exo* to *pol* direct switching, consistent with biochemical dataof T7 DNAp and single-molecule data of T4 DNAp [42, 117]. In a physiological setting it would be especially important to escape processive *exo* and continue replication since an error is quickly removed and further exonucleolytic activity would be deleterious. At the same time, we find no evidence for direct switching from *pol* to *exo*. Hence, our data shows that unbinding from *pol* and rebinding in *exo* is the main post-incorporation error-correction pathway for T7 DNAp. The bias for *exo* at higher tension has a clear analog in proofreading: when the PTS is destabilized by an error added in *pol*, a bias toward *exo* binding ensures that such errors are removed with high probability.
The biological function of the long pauses remains unclear. The length of the pauses indicates strong binding of the DNAp in the vicinity of the junction. Lack of tension-dependence of the rate into the long pause (Figure 4.4B), suggests a DNAp binds to ssDNA close to the junction. It has been noted before that DNAp has a strong affinity for ssDNA [42, 94]. At the replication fork, the recruitment of DNAp is facilitated by gp4; the helicase/primase is capable of binding up to three DNAp's, increasing the local DNAp concentration to directly start replication of the Okazaki fragment when the primer is synthesized [67, 112]. We speculate that the long pause state might represent inactive DNAp bound close to the junction, which could facilitate DNAp rebinding with its *pol* active site.
Recently, it was shown in a high-resolution optical tweezers study of bacteriophage Φ29 DNAp (a member of the *family B* polymerases) that a second polymerization state, with a slower, tension-dependent rate, is an obligatory intermediate to switch to proofreading activity [79]. In the case of T7 DNAp, we have found no evidence for such a polymerization state. The model for another *family B* DNA polymerase, T4 DNAp, was proposed to have a kinetic scheme that closely resembles the scheme presented here [117]. The kinetic model of T4 DNAp contains a paused state during polymerization and a direct transition from exo to pol. The paused intermediate is also present in our model. For T7 DNAp, the transition to this paused intermediate is dependent on tension, in contrast to the case of T4 DNAp. This particular discrepancy might arise from the specific assay performed. In the hairpin assay, DNA tension destabilizes the dsDNA ahead of DNAp. Contrarily, in our assay the tension is directly along the coordinate of DNA replication. Although the exact mechanisms of achieving high-fidelity replication might differ slightly for varying replicative DNA polymerases, they do seem to respond with remarkable similarities to DNA tension suggesting a similar mechanistic functioning of these polymerases.

Taken together, our findings indicate that the removal of errors by DNA polymerase should not be viewed as a precise control mechanism, but rather as a highly stochastic

process, prone to false positives, but tuned to avoid false negatives. Our data indicates that, at low tensions, 1 out of 5 events is exonucleolysis. Estimations indicate this would result in a cost of proofreading of up to ~10% of correctly incorporated nucleotides. Fersht et al. showed similar results in a study of the cost of proofreading by *E.coli* DNA polymerase III [48]. The energy difference between a correct and an incorrect nucleotide is small and in the range of the thermal energy. To, nonetheless, achieve the required high fidelity, DNAp needs to correct incorporated nucleotides frequently. The "paranoid" proofreading behavior of T7 DNA polymerase that we see evidence of, might represent a general alternative strategy employed by high-fidelity molecular machines to deal with thermal noise at the smallest scales of life [40, 106].

4.5 Methods

4.5.1 Trapping

A detailed description of the optical tweezers instrument can be found elsewhere [24,61]. In short, a Nd:YAG laser (3 W continuous wave, 1064 nm, Ventus 1064, Laser Quantum) was used to generate two optical traps. The laser beam was split into two orthogonally polarized beams using a polarizing beam splitter cube. The two beams were expanded with a 1:2.67 telescope system. One beam was steered by laterally displacing a telescope lens. Two traps were produced with a high-numerical aperture water-immersion objective. Two microspheres (1.87 μm, streptavidin-coated, Spherotech) were held in these traps, and we determined the distance between the two microspheres using a LabVIEW program (National Instruments), applying template-directed pattern matching. The displacement of the trapped microsphere with respect to the trap center was detected using a position-sensitive detector. Force and distance were simultaneously recorded by a Labview program, operating at a sampling rate of 25 Hz.

4.5.2 DNA construct

The pKYBI vector was restricted with KpnI and EcoRI (Fermentas). This resulted in a fragment of 8364 bp. The 5'-end overhang was filled in with biotin-14-dATP (Invitrogen) and dTTP (Fermentas) and Klenow exo$^-$ DNA polymerase (Fermentas). To the 3'-end, a 5'-biotinylated 29-mer (5'-cTcTcTcTctcttc tctcttctcttgtac-3', capitals indicating biotinylated nucleotides) was ligated, resulting in a DNA construct labeled with biotins on one strand, with a 25-nt 5'-end overhang. Therefore, exonucleolysis is first required to create ssDNA on which subsequently polymerization can take place. Switching between low and high tensions allows for repeated cycles of exonucleolysis on a single DNA molecule.

4.5.3 Fluidic system

Before starting experiments, the multichannel laminar flow cell was passivated with Roche Blocking Reagent (Roche Applied Sciences) to prevent non-specific binding of enzyme to the surface. DNA molecules were captured in buffer (10 mM Tris-HCl pH 7.7 and 50 mM NaCl) between two optically trapped microspheres (1.87 µm streptavidin-coated polystyrene beads, Spherotech) using a multichannel laminar flow cell. Tension to the DNA was applied by increasing the distance between the optical traps. Experiments were started by exchanging the buffer to buffer supplemented with 10 mM $MgCl_2$, T7 DNA polymerase (New England Biolabs) and dNTPs (0.6 mM each, Fermentas).

APPENDIX

Resolution

The spatial resolution in base pairs depends on the tension applied to the DNA. As tension on the DNA increases, the difference in length between ds- and ssDNA increases, resulting in an increased resolution. We have determined the resolution at 45 pN, by estimating the noise in the base pair over time signal. We used the spatial resolution at 45 pN as a reference to calculate the resolution as a function of tension by multiplication with changing length difference (Figure S4.1).

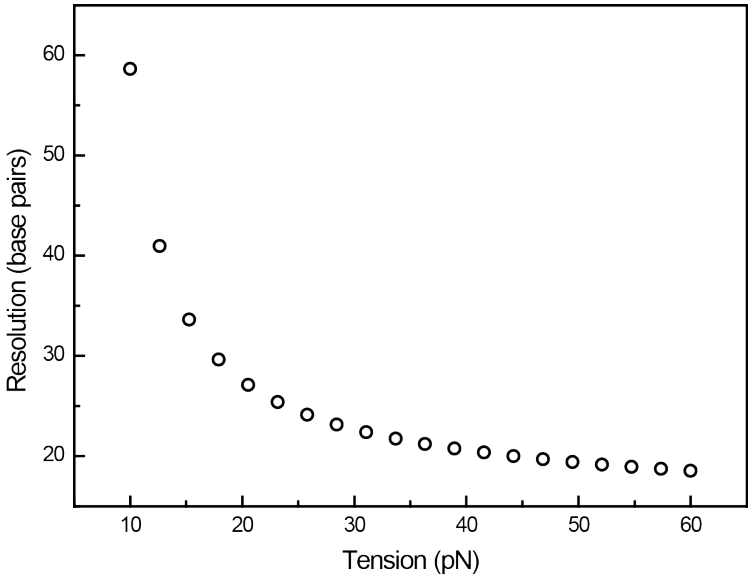

Figure S 4.1
Resolution in base pairs as a function of DNA tension. The resolution at 45 pN is 20 base pairs. The resolution at other tensions is calculated by multiplication with the relative lenth change from ds- to ssDNA. The resolution decreases as the tension decreases because the difference between ds- and ssDNA decreases.

Analysis of traces at constant tension

The extension of ds- and ssDNA differ for most tensions. Above ~6 pN the ssDNA is always longer while below that force it is shorter. We conduct all of our experiments above 6 pN. Thus, at constant tension, *pol* activity of DNAp will cause the DNA to become more dsDNA, resulting in a decreasing end-to-end length. In contrast, *exo* activity results in lengthening of the tether. Since the length of the DNA tether is known, the position at constant tension can be converted to the amount of base pairs and nucleotide. As a result the amount of incorporation and exonucleolysis can be given in nucleotides and followed over time. The obtained traces of conversion of nucleotides over time by DNAp were analyzed by finding the points at which the trend changes, i.e. where *pol* or *exo* starts or ends. To find these break points, a sliding window was fitted with two connected linear fits. The point at which these lines connect was varied throughout the window to find the break point that gave the smallest sum of residues. Each window therefore yielded a break point. Further analysis was needed to separate break points from artifacts. Selection was done on the occurrence and goodness of fit of the found break points. In the end, a linear fit to the data between the breakpoints was used in our analysis of the kinetic mechanism of DNAp. Linear fits with rates below 25 nt s^{-1} were marked as pauses. As a processivity threshold the spatial resolution given in figure S1 was used. Events shorter than 0.4 seconds were discarded.

Polymerization and exonucleolysis rates

The binned histograms of the rates of polymerization and exonucleolysis for all different tensions are shown in Figure S4.2.

Off rates

The durations of *pol* and *exo* activities were fitted with single exponential functions. The resulting fits of both activities at different tensions are shown in Figure S4.3.

Fraction of long pauses

The fraction of long pauses (of the total amount of pauses) does not significantly depend on the DNAp-concentration. At 15 nM the fraction is 0.25 ± 0.03 (mean ± SD), at 25 nM this fraction is 0.33 ± 0.04 after 1000 bootstraps. This suggests that the rate of going into a long pause scales with DNAp-concentration, similar to the rates into the *pol* and *exo* active sites. As explained in Section 4.3, the duration of a long pause is dominated by the slow return to the solution state.

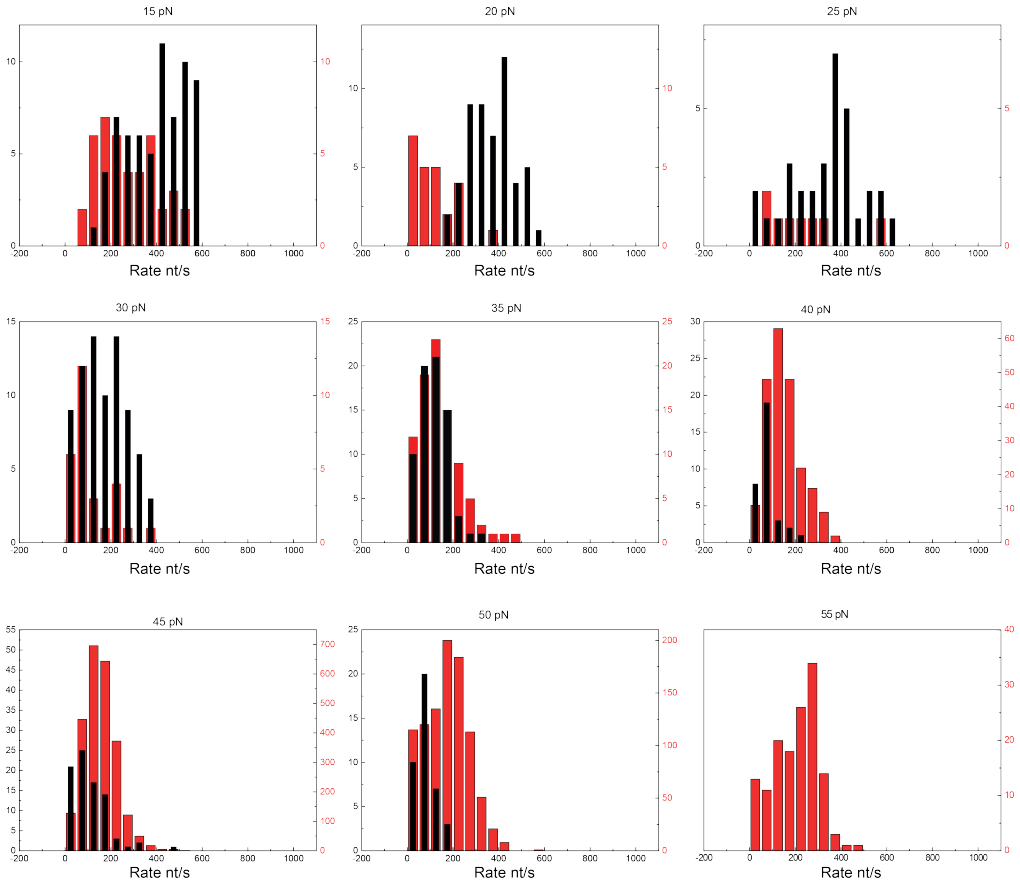

Figure S 4.2
Histograms of the rates of polymerization (black) and exonucleolysis (red) for different tensions.

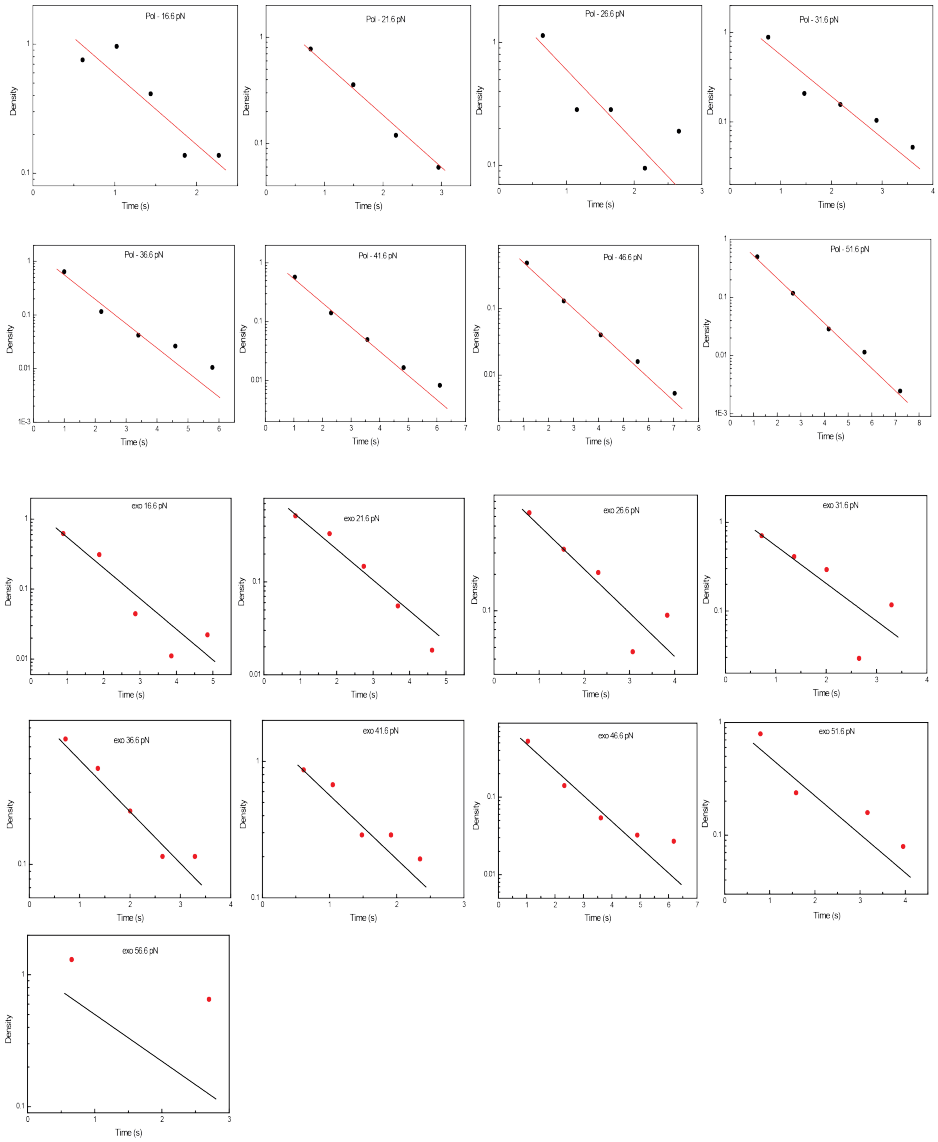

Figure S 4.3
Off rates of both *pol* and *exo*. For all tensions the length of *pol* and *exo* events are described as a single exponential. The pause durations are log-binned.

Net replication rate

In previous studies the reported rates of polymerization and exonucleolysis probed at different tension; below 35 pN there was tension-dependent polymerization; as the tension increased further, exonucleolysis started to occur with no or only small dependency on tension [79,180]. We observe both *pol* and *exo* over the full range of applied tensions (15-55 pN). However, we are able to reconstruct the previously reported behavior by weighing together both rates with respect to their relative occurrences (Figure 4.2C). We do not take the off rates of both activities into account since they are similar for all tensions. Pauses are not taken into account, as their duration is dependent on the concentration of DNAp used.

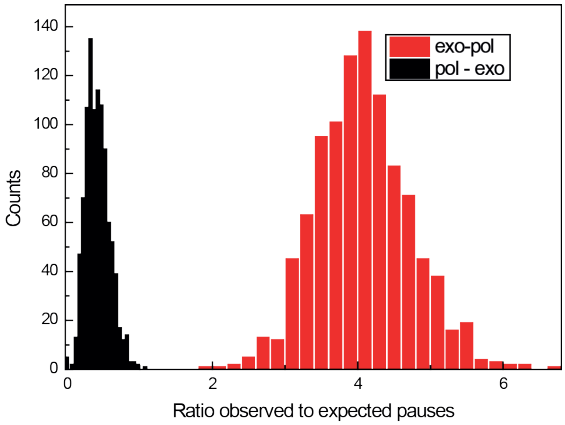

Figure S 4.4
Ratio of observed sub-resolution pauses to expected sub-resolution pauses for *pol-exo* (black) and *exo-pol* (red). The histogram is a result of 1000 bootstraps.

Direct switches

In our assay we cannot resolve pauses below 0.4 seconds. However, direct switches (switches from *pol* to *exo* or vice versa) could still be resolved. The different directions of *pol* and *exo* in our traces results in a clear, distinguishable trend change. Figure 4.1D shows all possible trend changes. As long as the activities are above the threshold direct switches would still be observed. Figure S4.4 shows the overrepresentation of *exo* to *pol* switches over the expected amount of sub-resolution pauses, indicating a direct switch. In contrast, the switching from *pol* to *exo* is observed rarely and therefore attributed to unresolved pauses.

Modeling and Maximum likelihood fitting results

We use continuum-time random walk theory to determine the four probability distributions for the time it takes to reactivation into *exo* or *pol* when having entered the pause from *exo* or *pol*. The results are

$$
\begin{aligned}
P_{EE} &= \frac{\boldsymbol{k}_{ES}}{\boldsymbol{k}_{EP}+\boldsymbol{k}_{ES}}\frac{\boldsymbol{k}_{SE}}{\boldsymbol{k}_{SE}+\boldsymbol{k}_{SP}}P_{sol}(t)\\
P_{EP} &= \frac{\boldsymbol{k}_{ES}}{\boldsymbol{k}_{EP}+\boldsymbol{k}_{ES}}\frac{\boldsymbol{k}_{SP}}{\boldsymbol{k}_{SE}+\boldsymbol{k}_{SP}}P_{sol}(t)+\frac{\boldsymbol{k}_{EP}}{\boldsymbol{k}_{EP}+\boldsymbol{k}_{ES}}\delta(t)\\
P_{PE} &= \frac{\boldsymbol{k}_{PS}}{\boldsymbol{k}_{PS}+\boldsymbol{k}_{PI}}\frac{\boldsymbol{k}_{SE}}{\boldsymbol{k}_{SE}+\boldsymbol{k}_{SP}}P_{sol}(t)\\
P_{PP} &= \frac{\boldsymbol{k}_{PS}}{\boldsymbol{k}_{PS}+\boldsymbol{k}_{PI}}\frac{\boldsymbol{k}_{SP}}{\boldsymbol{k}_{SE}+\boldsymbol{k}_{SP}}P_{sol}(t)+\frac{\boldsymbol{k}_{PI}\boldsymbol{k}_{IP}}{\boldsymbol{k}_{PS}+\boldsymbol{k}_{PI}}e^{-\boldsymbol{k}_{IP}t}
\end{aligned}
$$

where the notation for the rates are as indicated in Figure 4.4, and the subscript on the probability densities correspond starting in *exo/pol* (E/P) followed by ending in *exo/pol* (E/P). Further, the solution escape distribution is given by

$$P_{sol}(t) = \frac{\boldsymbol{k}_{SE}+\boldsymbol{k}_{SP}}{\omega_+-\omega_-}\left(e^{-\omega_+ t}(\boldsymbol{k}_{LS}-\omega_+)\right)$$

with the characteristic rates

$$\omega_\pm = \frac{1}{2}\left(\boldsymbol{k}_{LS}+\boldsymbol{k}_{SL}+\boldsymbol{k}_{SE}+\boldsymbol{k}_{SP} \pm \sqrt{(\boldsymbol{k}_{LS}+\boldsymbol{k}_{SL}+\boldsymbol{k}_{SE}+\boldsymbol{k}_{SP})^2-4\boldsymbol{k}_{LS}(\boldsymbol{k}_{SE}+\boldsymbol{k}_{SP})}\right)$$

From these distributions we construct the maximum likelihood function

$$ML(\boldsymbol{k}_{XY}) = -\sum_{X\epsilon\{E,P\}}\sum_{Y\epsilon\{E,P\}}\sum_{i=1}^{I^{XY}} lnP_{XY}(t_i^{XY})$$

where the sums run over the observed pauses of each type, with experimentally measured pause durations. The rates are now estimated by numerically minimizing ML over all the rates of the model (see Figure 4.4A). The errors are calculated by bootstrapping the data 300 times and we report the one-sigma confidence interval.

Distances to transition barrier

Assuming that the force dependence of any of the rates is dominated by one free-energy barrier, it possible to describe the force dependence in terms of a distance to a transition state:

$$\boldsymbol{k}_{XY} = \boldsymbol{k}^0_{XY} e^{-\delta_{XY} f / k_B T}$$

This assumption is not always justified, which is the reason for not using it directly in our fits. Instead, we have binned the forces in three bins to be able to at least qualitatively judge the deviation from this basic form (which would result in straight lines in the log-plots of Figure 4.4). Within the errors, all but the rate $\boldsymbol{k}_{SP}$ seem to be described by the above form. For $\boldsymbol{k}_{SP}$ we allow for next order variations with respect to force, according to

$$\boldsymbol{k}_{SP} = \boldsymbol{k}^0_{SP} e^{-(\delta_{SP} f + \gamma_{SP}(f-\bar{f})^2 - \gamma_{SP}\bar{f}^2)/k_B T}$$

Here we have introduced the average force $\bar{f}$ to keep the interpretation of δ_{SP} as the distance to the transition state in the center of the probed force range. The corresponding distances to the transition states and zero-force rates are given by

Table S 4.1
Rates at zero tension and distance to the barriers for all fitted rates.

	$\boldsymbol{k}_{ES}$	$\boldsymbol{k}_{EP}$	$\boldsymbol{k}_{IP}$	$\boldsymbol{k}_{LS}$	$\boldsymbol{k}_{PI}$	$\boldsymbol{k}_{PS}$	$\boldsymbol{k}_{SP}$	$\boldsymbol{k}_{SE}$	$\boldsymbol{k}_{SL}$
$\boldsymbol{k}^0$	1.15	0.16	30.63	0.02	1.65	1.42	0.06	0.05	0.04
δ	0.02	0.19	0.42	-0.13	0.16	0.08	0.47	-0.15	-0.05

The distance to the energy barrier for the different fitted rates. In Table S4.1, we give the zero-tension rate for all fitted rates ($\boldsymbol{k}_0$). Since the rates are extrapolated well beyond the probed tension regime, caution should be taken interpreting $\boldsymbol{k}_0$.

Cost of proofreading

For above reasons on the extrapolation of the fitted rates, we use data at 15 pN to estimate the cost of proofreading at zero tension. At 15 pN, $\boldsymbol{k}_{pol}$ is ∼500 nt s^{-1} and $\boldsymbol{k}_{exo}$ is ∼150 nt s^{-1} (Figure 4.2A), with the duration of both activities for ∼1 sec (inset Figure 4.2A). The probability of exonucleolysis at 15 pN is 0.2 (Figure 4.2B). This would correspond to the removal of approximately 150 nt for every 2000 nucleotides incorporated, and a cost of 7.5% for proofreading.

Alternative Models

Model A

Several alternative models that could describe the data were considered. In this section we explicitly discuss three seemingly natural alternative models and their limitations. The first model is one with four different states that are all interconnected (polymerization, exonucleolysis, unbound in solution and the long paused state) (Figure S4.5A). Two reasons for discarding this model reside in the memory effect and the entrance probability of the long paused state. The memory effect in this model would be regulated through the long paused state, which would have to bias the binding at high tensions into the polymerization state. We have found, however, that the access (and exit) of long pauses does not depend on the activity before the pause. Therefore the long pauses state could not bias the binding into the polymerization state and describe the observed memory effect. Since the fractions of paused DNAp entering the long paused state are similar coming from or going into *pol* and *exo*, there is probably one common entry and exit from the long paused state.

Model B

The second model that we discuss explains the memory effect not through a different state of the enzyme, but different states of the DNA (Figure S4.5B). In this model, two unbound states exist, one in which the primer-template structure is aligned (S1) and one in which the PTS is frayed and several base pairs are opened up (S2). This model would be able to describe the data if the rates between the non-frayed and frayed DNA state are force-dependent. The effect of force could only be described, however, if the rates between the two states are on the order of seconds. In data previously acquired in our lab on the melting of DNA in the overstretching transition it was found that the melting and reannealing of tens of base pairs occurs multiple times per second [62]. At this rate the fraying of a few base pairs at tensions below the overstretching transition, would equilibrate quickly and the force dependency would be lost. For that reason, this model was discarded.

Model C

The last model we discuss here is similar to the model presented in the main text (Figure S4.5C). In this case, an extra paused state is added to exonucleolysis, to justify the memory effect. This model, however, is not able to simultaneously capture the observed comparable amounts of *exo-pol* and *pol-pol* transitions at low forces, and the increased *pol-pol* transitions at high forces.

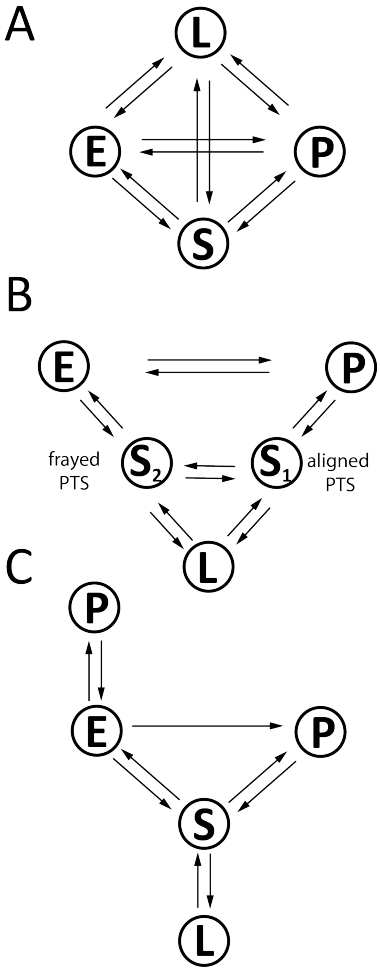

Figure S 4.5
Alternative models. See text in Appendix for discussion.